Paudyal, P., Yan, S., & Simeonidou, D. (2020). Toward Deployment of ML in Optical Networks, Transfer Learning, Monitoring and Modelling. In *Asia Communications and Photonics Conference/International Conference on Information Photonics and Optical Communications 2020 (ACP/IPOC)* (pp. S4C.2). Optical Society of America (OSA).

Toward Deployment of ML in Optical Networks, Transfer Learning, Monitoring and Modelling

Paurakh Paudyal[1], Shuangyi Yan[1],and Dimitra Simeonidou[1]

[1] HPN group, Smart Internet Lab, University of Bristol shuangyi.yan@bristol.ac.uk

Abstract *We present a novel approach for Quality of Transmission estimation using hybrid modelling and transfer-learning. Our method reduces the training data requirement by 80% while obtaining an MSE of 0.27dB. The approach facilitates a streamlined ML life-cycle for data collection, training and deployment.*

Introduction

The emerging 5G networks designed to support high-capacity network applications will bring an unprecedented amount of dynamic traffic to the underlying optical network infrastructures[1]. Future optical networks will need to be evolved to be more dynamic, with the ability to establish network connections with reduced margins to improve hardware utilisation[2]. Consequently, precise information about the quality of transmission (QoT) of the unestablished light paths as well as the impact of newly established light paths on the previous channels is of vital importance to operating low-margin optical networks efficiently.

With the unparalleled combination of high accuracy and low computational complexity in inference, Machine Learning (ML) based approaches have been explored to provide promising solutions in QoT estimation with either synthetic data[3] or pre-collected network operation data[4]. These solutions, based on artificial neural networks (ANN), face big challenges in scalability as training and inferencing of ML models are carried out on the same network. Recently, transfer learning (TL)[5] has been used to solve the scalability of the ANN-based QoT prediction[6]–[8] in optical networks. However, there are still several challenges that need to be addressed before their widespread deployment can become a reality. The large amounts of data that are required for ML model training are not yet available, particularly during the early phase of the fibre life-cycle when network monitoring data is lacking. Additionally, ML models must be responsive to system changes caused by component wear and ageing, and constantly evaluate their own efficacy so as to prevent inadequate quality of service. A coherent life-cycle for ML models is required to formalise the process of designing, testing, and deploying ML models on optical networks.

In this paper, we propose a streamlined ML life-cycle for optical networks which utilises TL to combine synthetic data and practical network observation data. Synthetic data gathered through coarse analytical modelling is used to obtain a QoT-prediction model with acceptable precision in the absence of practical network data. The QoT prediction model is then retrained and fine-tuned to obtain high precision with practical data. The performance of the TL assisted ANN is evaluated by comparing it to a baseline ANN trained from scratch. The TL-assisted ANN achieves an MSE of 0.267 dB, equal to that of the baseline ANN, despite being trained on only 20% (200 samples) of the practical data used to train the baseline model. The training time is reduced for the TL assisted ANN, taking 6.67s in comparison to 19.47s for the baseline ANN. Our proposed approach reduces the volume of practical data required to train an ANN for QoT prediction, and facilitates rapid training and deployment of these predictors in future commercial optical networks.

ML Life-cycle towards deployment in optical networks

Our proposed four-phase approach for the ML life-cycle is outlined in Figure 1. Transfer learning is used to connect available knowledge and practical network status. First, a source learner is trained on synthetic data gathered offline through available modelling or simulation tools. The design phase consists of model selection, training and optimisation. Initial values for hyper-parameters are set arbitrarily and optimised by performing a Grid Search across many hyper-parameter combinations. Then, the parameters of the trained source learner are transferred to the target (TL assisted) learner, which is fine-tuned with practical monitoring data from the optical network. This approach achieves convergence faster than training from scratch while reducing the re-

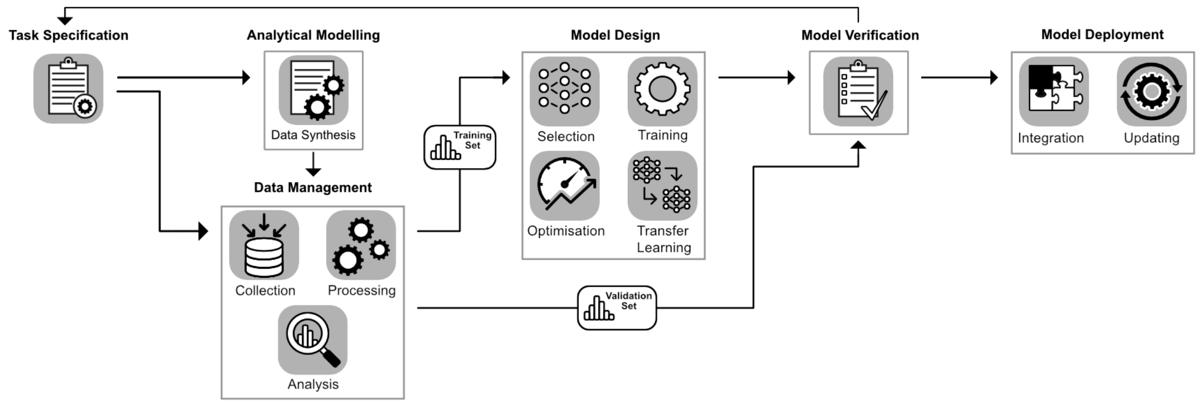

Fig. 1: Hybrid model life-cycle management for machine learning algorithms in optical networks

quired amount of training and validation data. The third phase is model validation, which aims to evaluate the model's ability to generalise for new, unseen network operations data. The results of this stage determine margins that must be utilised during deployment to ensure adequate quality of service. Successful models can be integrated into the software defined networking (SDN) controller for autonomous and dynamic light-path allocation. A remodelling algorithm accounting for all potential sources of system changes (e.g. component wear and ageing) can be implemented with practical network validation. With the proposed ML model management, network operation scenarios with complex topology and dynamic channel configurations can be offered ML models for QoT predictions. In the following section, we demonstrate the first three phases for ML model provision.

Transfer Learning Based QoT Estimator

We utilise TL to reduce the required training data and increase speed of convergence for our ANN model. A source domain, $\mathcal{D}_S$, and source task, $\mathcal{T}_S$ are defined. $\mathcal{D}_S$ is comprised of the feature space $\mathcal{X}_S$ and a marginal probability distribution $P_S(x)$. $\mathcal{T}_S$ is comprised of the target space, $\mathcal{Y}_S$ and the predictive function $f(\cdot)$. Similarly, a target domain, $\mathcal{D}_T$, and target task, $\mathcal{T}_T$ are also defined. For our purposes, $\mathcal{D}_S$ is the synthetic data set and $\mathcal{D}_T$ is the practical network data. TL aims to improve the learning of the target predictive function $f(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$ but $\mathcal{T}_S = \mathcal{T}_T$ (both tasks are OSNR estimation). We utilise parameter transfer[7], and transfer a certain amount of parameters (weights, biases etc.) from the source learner to the target learner. This can be explained mathematically by:

$$w_S = w_0 + v_S \; and \; w_T = w_0 + v_T \qquad (1)$$

where w_S and w_T are parameters of the ANNs used for the source task and the target task respectively. w_0 is the set of parameters shared between both tasks, while v_S and v_T are task specific parameters. The cost function for the target model can then be written:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i; w_0, v_T))^2 \qquad (2)$$

Where y_i is the actual output for the i^th data point and $f(x_i)$ is the output predicted by the target learner. Hence parameter transfer induces faster training of the target model by making w_0 known. The two step process therefore is comprised of the initial training of the source learner ANN on synthetic data, followed by the transferring of parameter knowledge to the target learner which is fine-tuned to practical network data.

We built on our previous work in[4] and used the ANN architecture as a starting point for the source learner. Synthetic data was gathered on a simulated version of the national dark fibre facility (NDFF) network using route planning software and the Gaussian noise (GN) analytical model[9]. The network topology can be seen in Figure 2a. The training and test data-sets consisted of 9000 and 1000 data points respectively. The feature space was comprised of launch power, EDFA input/ output power pairs for every node, and the channel coding vector resulting in a 43 dimensional feature space. There are 8 outputs corresponding to the total OSNR for each of the channels at the end of the link. Channel coding is achieved through hot coding, where a vector of [1 1 0 0 1 0 0 0] represents channel 1,2 and 5 being switched on. The ANN can thus infer the significance of a 0 in the channel vector, and predict a 0 for the OSNR of that channel. Similarly, the EDFA input/ output powers behave as a hot

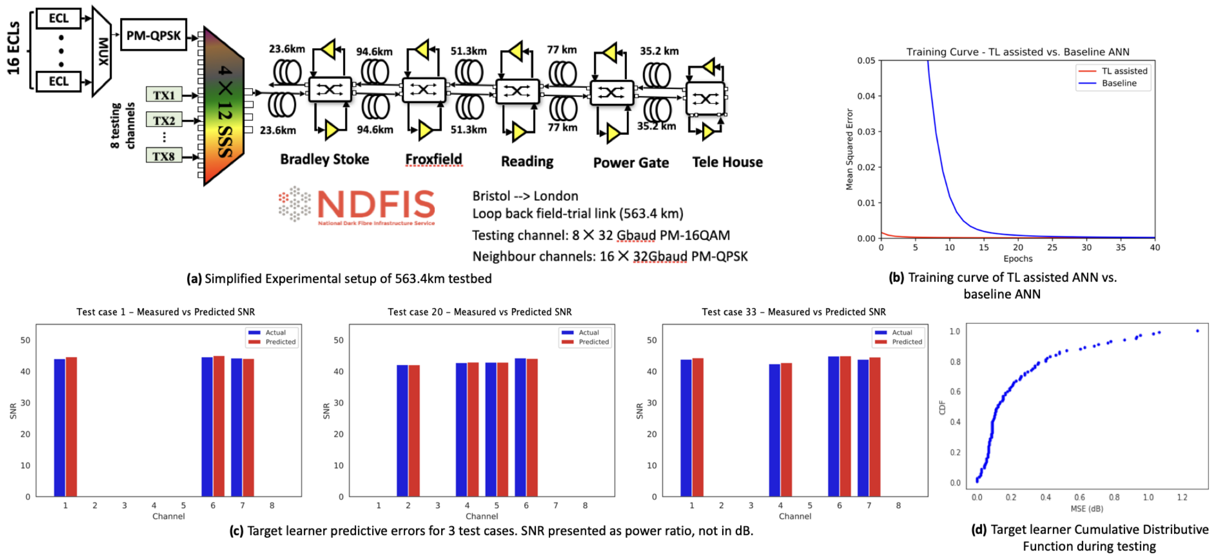

Fig. 2: Experimental setup and results

coded route vector by the inclusion of 0 values for all EDFAs that were not crossed on a given path, allowing the ANN to make inferences based on the route taken. We used a rectified linear unit (ReLU) function $f(x) = max(0, x)$ as the activation for all neurons except those in the output layer, where the sigmoid function was used. An Adaptive Moment Estimation (Adam) was chosen as a stochastic gradient descent algorithm. The architecture of the ANN with two hidden layers was [43,40,20,8]. Batch size and number of epochs were 16 and 150 respectively. The hyper-parameters and architecture of the source learner are transferred to the target learner, which is re-trained on 200 random samples of practical data. After testing multiple parameter transfer schemes, fine-tuning (all parameters transferred) was chosen as the best approach, indicating heavy correlation between $\mathcal{T}_S$ and $\mathcal{T}_T$.

Results and Discussion

The performance of the target learner was evaluated against a baseline model trained from scratch with 500% more practical data (1000 samples). The training curves for the target learner and baseline model can be seen in Figure 2b. The target learner obtained convergence quicker than the baseline model, requiring only 2 epochs as opposed to 24. Training time was also reduced from 19.47s for the baseline model to 6.67s for the target learner. Both models achieved an MSE of 0.27dB across 100 test data points and a 90th percentile accuracy of 0.704dB. The performance of the target learner for 3 test cases is shown in Figure 2c, and a plot of the cumulative distributive function (CDF) can be seen in Figure 2d. Our results indicate that robust QoT predictors can be trained with far less practical data than previously thought necessary. The proposed ML life-cycle facilitates the rapid training and deployment of ML based QoT predictors in dynamic optical networks.

Conclusion

In this paper, a streamlined ML life-cycle is presented for future deployment in commercial optical networks. Based on the life-cycle, a TL-based QoT estimation is implemented with reduced training data requirements and faster training times. Our approach paves the way for large-scale, rapid deployment of QoT predictors with complex network operation scenarios, to fully support dynamic optical networks during the 5G era.

Acknowledgements

This work was supported by UK EPSRC through the TOUCAN project (EP/L020009/1) and the NDFF project (EP/S028854/1).

References

[1] M. Fiorani et al, in *ANTS*, 2014, pp. 1–6.

[2] Y. Pointurier, *IEEE/OSA JOCN*, vol. 9, no. 1, A9–A17, Jan. 2017.

[3] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, *IEEE/OSA JOCN*, vol. 10, no. 2, A286–A297, Feb. 2018.

[4] Z. Gao, S. Yan, J. Zhang, M. Mascarenhas, R. Nejabati, Y. Ji, and D. Simeonidou, *JLT*, vol. 38, no. 9, pp. 2646–2655, 2020.

[5] F. Zhuang et al, *arXiv preprint arXiv:1911.02685*, 2019.

[6] W. Mo et al, in *OFC 2018*, IEEE, 2018, pp. 1–3.

[7] L. Xia et al., *Optics Express*, vol. 27, no. 14, p. 19 398, Jul. 8, 2019.

[8] D. Azzimonti et al, in *OFC*, San Diego, California: OSA, 2020, M4E.1.

[9] P. Poggiolini, *JLT*, vol. 30, no. 24, pp. 3857–3879, 2012.